HHS Public Access

Author manuscript

Neuron. Author manuscript; available in PMC 2015 May 12.

Published in final edited form as:

Neuron. 2012 May 24; 74(4): 603–608. doi:10.1016/j.neuron.2012.05.001.

Data visualization in the neurosciences: overcoming the curse of dimensionality

Elena A. Allen[*,1,2], **Erik B. Erhardt**[1,3], and **Vince D. Calhoun**[1,4]

[1]The Mind Research Network, 1101 Yale Blvd. NE, Albuquerque NM 87106, USA

[2]Departments of Biological and Medical Psychology and the K.G. Jebsen Center for Research on Neuropsychiatric Disorders, University of Bergen, 5009 Bergen, Norway

[3]Department of Mathematics and Statistics, University of New Mexico, Albuquerque NM 87131, USA

[4]Departments of Electrical and Computer Engineering, Psychiatry, and Neuroscience, University of New Mexico, Albuquerque NM 87131, USA

Abstract

In publications, presentations, and popular media, scientific results are predominantly communicated through graphs. But are these figures clear and honest, or misleading? We examine current practices in data visualization and discuss improvements, advocating design choices which reveal data rather than hide it.

Visualizations are vital tools for neuroscientists of every discipline, affording the ability to reveal relationships in large datasets and communicate information to a broad audience. But with the great power of graphs, one might say, comes great responsibility. Graphs can be fundamentally misleading about underlying data, and design choices can skew viewers' perceptions leading them toward incorrect conclusions (Jones, 2006). For example, recent studies suggest that results rendered on aesthetically pleasing brain images are perceived as more persuasive and credible than identical information presented in other formats (Keehner et al., 2011; McCabe and Castel, 2008). Beyond the attractiveness of displays, readers may also be misled by the frequent errors that plague scientific figures (Cleveland, 1984) or a lack of sufficient information. In the words of statistician and graphic design expert Howard Wainer, effective data visualization must "remind us that the data being displayed do contain some uncertainty" and "characterize the size of that uncertainty as it pertains to the inferences we have in mind" (Wainer, 1996). It is our impression that such descriptions (along with more basic elements) are often lacking from published figures. In this NeuroView, we perform a survey of figures from leading neuroscience journals with an eye towards clarity and the portrayal of uncertainty. Based on survey results, we discuss methods to improve graphics (particularly for large datasets where visualization poses a challenge) and propose a set of figure guidelines in the form of a checklist (Table 1). We hope these recommendations, compiled from a number of excellent resources on data

[*]Corresponding author: eallen@mrn.org.

visualization (Lane and Sándor, 2009; Tufte, 2001; Wainer, 1996) may be used by both internal and external reviewers to help evaluate figures for clarity and completeness.

Surveying the Field

We sampled 288 articles published in 2010 from 6 neuroscience journals (*Frontiers in Systems Neuroscience*, *Human Brain Mapping*, *Journal of Neuroscience*, *Nature Neuroscience*, *NeuroImage*, and *Neuron*) and examined the 1451 figures therein. We surveyed four basic features that were applicable to nearly all graphs and addressed Wainer's points above. Survey questions were: 1) Is the dependent variable or quantity of interest labeled? 2) Is the scale of the dependent variable indicated? 3) Where applicable, is a measure of uncertainty displayed? 4) Is the type of uncertainty (e.g., standard error bars or confidence intervals) defined in the figure or accompanying legend? Examples of these graphical features are shown in Figure 1A for 2-dimensional (2D) and 3D datasets.

Survey results, shown in Figure 1B, overwhelmingly suggest that graphical displays become less informative as the dimensions and complexity of datasets increase. Compared to graphs of 2D data, 3D displays provide poorer descriptions of the outcome of interest and rarely provide an indication of uncertainty. Only 43% of 3D graphics label the dependent variable (meaning that if you were asked "what is being plotted here?" you would be able to answer less than half of the time) and only 20% portray the uncertainty of reported effects. Even for 2D data, the proportion of graphs displaying uncertainty is lower when explanatory variables are continuous (and typically take on many values) than when they are categorical (and typically represent a few conditions; Fig. 1C). Of 2D figures that *do* indicate uncertainty, nearly 30% fail to define the type of uncertainty or variability being portrayed. Given the plurality of interpretations connoted by an error bar (e.g., a standard deviation [s.d.] of the sample, a standard error of the mean [s.e.m.], a range, a parametric confidence interval [CI] of the mean, a bootstrap CI, a Bayesian probability interval, a prediction interval, etc.), it is unclear how including it without a proper label would offer readers any further understanding of the data; in contrast, the poor labeling or omission of error bars has been shown to encourage misinterpretation (Cumming and Finch, 2005; Vaux, 2004; Wainer, 1996).

A breakdown of results by journal (see supplementary analysis at mialab.mrn.org/datavis) further highlights the issue of data dimensionality in visualization: journals with lower proportions of 2D and 3D graphical features are those that primarily publish neuroimaging and systems-level findings, where results are often distilled from very large datasets using a hierarchy of models. That the so-called "curse of dimensionality" extends to the realm of data visualization is not surprising. Dependent variables are more difficult to label when they represent abstract parameter estimates rather than directly-measured quantities; uncertainty is more challenging to render when datasets require error surfaces rather than error bars. However, these results are undesirable. As datasets become more complex, displays should become increasingly informative, elucidating relationships that would be inaccessible from tables or summary statistics. In the next section, we provide examples of creating more informative displays for simple and complex datasets by making design choices that reveal data, rather than hide it.

Show More, Hide Less

Consider a simple experiment where a researcher investigates the effect of different conditions on a single response variable. Having collected 50 samples of the response variable under each condition A, B, and C, how should the researcher visualize the data to best inform themselves and their audience of the results? Figure 2 provides three possible designs. In panel I, a bar plot displays the sample mean and s.e.m. under each condition. With no distributional information provided, the data-density is quite low and the same information could be provided in a single sentence, e.g., "mean response ± s.e.m for conditions A, B, and C were 4.9 ± 0.4, 5.0 ± 0.4, and 5.2 ± 0.4, respectively". Panel II offers some improvement with box plots displaying the range and quartiles of each sample. This design reveals that response variables may take on both positive and negative values (hidden in panel I), and that condition B may be right-skewed. Distributional differences are better understood in panel III when using violin plots to display kernel density estimates (smoothed histograms) of each dataset (Hintze and Nelson, 1998). Violin plots make the skew in condition B more apparent and reveal that responses in condition C are bimodal (hidden in panels I and II). Although the additional distributional information in panel III does not change our initial inference that sample means are similar between conditions, we are certainly not likely to make the misinterpretation that condition has no effect on the response. Distributional differences also suggest that assumptions of the ANOVA (or other parametric models) may not be met, and that the mean may not be the most interesting quantity to investigate.

This example is not meant to imply that bar plots should always be avoided in favor of more complex designs. Bar plots have numerous merits: they are easy to generate, straightforward to comprehend, and can efficiently contrast a large number of conditions in a small space. They are particularly effective for displaying frequencies or proportions (as in Fig. 1), where binary data samples are transformed into a height that intuitively reflects the fraction of "successes". Yet bar plots are also commonly used in scenarios where the distance from zero is not meaningful, and where distributional information would be of great benefit to readers. In roughly the same amount of space required by a bar plot, one can portray the full shape of distributions and overlay descriptive statistics, inferential statistics related to hypothesis testing, or even individual data points, creating a so-called "bean plot" (Kampstra, 2008). By increasing the amount of information available to the viewers, we allow them to assess the appropriateness of related statistical analyses and make their own inferences.

In Figure 3, we apply the guiding principle to "show more, hide less" to high-dimensional electroencephalographic (EEG) and functional magnetic resonance imaging (fMRI) datasets. We portray the results using a common design (panel I) and a modified design (panel II), where each change is arrived at by following the guidelines in Table 1.

Figure 3A presents data from an EEG visual flanker task. Subjects were asked to indicate the direction of a visual target which appeared shortly after the presentation of flanking distracters. For each participant, multi-channel EEG timeseries were decomposed using independent component analysis and a single component best matching the expected fronto-central topography for a performance monitoring process was selected for further analysis

(Eichele et al., 2010). Here, we ask how the extracted event-related potential (ERP) differs according to the subject's response (i.e., correct or incorrect). Panel I provides a typical portrayal of results, where mean ERPs are displayed for each condition. As Table 1 recommends, the axes are labeled, variable units are indicated, and experimental conditions are distinguished by line color with direct annotation on the plot. While this panel is clear, it is not complete: there is no portrayal of uncertainty. In panel II, we add 95% confidence bands around the average ERPs. The confidence bands are made slightly transparent to highlight overlap between conditions and to maintain the visual prominence of the means. Confidence intervals clarify that there is greater uncertainty in the error response than the correct response (since subjects make few errors), and that there is insufficient evidence to conclude a response difference after ~800 ms. In panel II we also add verbal descriptions and additional annotation to the graphic (Lane and Sándor, 2009; Tufte, 2001). Labels indicate that the timeline is relative to the presentation of the target stimulus, and specify our null and alternative hypotheses as well as the alpha level (Type I error rate) chosen to determine statistical significance. Integrating descriptions into the figure (rather than the legend) discourages misinterpretation and permits readers to understand the display more quickly. Of course, annotation must be used judiciously and should not overwhelm or detract from the data visualization itself.

Figure 3B portrays results from an auditory oddball event-related fMRI experiment. Participants responded to target tones presented within a series of standard tones and novel sounds. Blood oxygenation level-dependent (BOLD) timeseries at each brain voxel were regressed onto activation models for the target, novel, and standard stimuli (Kiehl et al., 2001). Here, we ask what brain regions might be involved in the novelty processing of auditory stimuli and compare beta parameters between novel and standard conditions. Panel I presents voxelwise differences between beta coefficients using a widely reproduced design: functional-imaging results are thresholded based on statistical significance and overlaid on a high-resolution structural image. Following Table 1, the variable of interest is labeled, the colormap is sensible for the data and is mapped with symmetric endpoints, and annotation clearly indicates the directionality of the contrast (i.e., "Novel – Standard). This design provides excellent spatial localization for functional effects, but is not without problems. The display does not portray uncertainty and has a remarkably low data-ink ratio due to the prominent (non-data) structural image and sparsity of actual data (Habeck and Moeller, 2011). More crucially, the design encourages authors to hide results not passing a somewhat arbitrary statistical threshold. Given numerous correction methods and little consensus on the appropriate family-wise Type I error rate (Lieberman and Cunningham, 2009) authors may arrive at a "convenient" threshold to reveal visually appealing and easily explained results. This design reduces a rich and complex dataset to little more than a dichotomous representation (i.e., "significant or not?") that suffers from all the limitations of all-or-none hypothesis testing (Harlow et al., 1997) .

Rather than threshold results, we suggest a dual-coding approach to represent uncertainty (Hengl, 2003). As shown in panel II, differences in beta estimates are mapped to color hue, and associated paired *t*-statistics (providing a measure of uncertainty) are mapped to color transparency. Compared to panel I, no information is lost. Transparency is sufficient to

determine structural boundaries and statistical significance is indicated with contours. However substantial information is gained. The quality of the data is now apparent: large and consistent differences in betas are wholly localized to gray matter, while white matter and ventricular regions exhibit very small or very uncertain differences. In addition, isolated blobs of differential activation in panel I are now seen as the peaks of larger contiguous activations (often with bilateral homologues) that failed to meet significance criteria. The modified display also reveals regions in lateral parietal cortex, medial prefrontal cortex, and posterior cingulate cortex with reduced activation to novel stimuli compared to standard tones. These brain areas coincide with the so-called “default-mode network”, a system preferentially active when subjects engage in internal rather than external processes (Buckner et al., 2008). We hope to impress upon the reader the wealth of findings that can be revealed simply by un-hiding data. To encourage the use of this approach we provide sample MATLAB scripts for hue and transparency coding on our website (mialab.mrn.org/datavis).

Along with increased annotation, panel II also displays the beta parameters for individual subjects, averaged over clusters of voxels passing significance (Fig. 3B1,B2). The 2D plots remove dependence on color-mapping (which is more difficult for viewers to decode than position along an axis (Cleveland and McGill, 1985)) and allow us to access the data in greater detail. Scatter plots indicate the beta estimates for each condition (rather than just the difference), reveal the degree of variability across subjects (and the absence of outliers), and validate our “paired” statistical approach, since beta values covary across conditions.

Concluding Remarks

A single figure may portray experimental data painstakingly collected over months or even years. Rather than use standard designs such as bar plots and thresholded maps that hide these data, we, as authors, peer-reviewers, and editors, can establish new standards for visualizations that reveal data and inform readers.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Christian Habeck & James Moeller, whose commentary helped to motivate this work, Tom Eichele, for his contribution of the EEG data, and Kent Kiehl & Godfrey Pearlson for their contribution of the fMRI data. We also thank Christian Habeck and Tom Eichele for valuable discussions throughout the completion of this work.

References

Buckner RL, Andrews-Hanna JR, Schacter DL. The brain's default network: Anatomy, function, and relevance to disease. Annals of the New York Academy of Sciences. 2008; 1124:1–38. [PubMed: 18400922]

Cleveland WS. Graphs in scientific publications. American Statistician. 1984; 38:261–269.

Cleveland WS, McGill R. Graphical perception and graphical methods for analyzing scientific data. Science. 1985; 229:828–833. [PubMed: 17777913]

Cumming G, Finch S. Inference by eye: confidence intervals and how to read pictures of data. American Psychologist. 2005; 60:170–180. [PubMed: 15740449]

Eichele H, Juvodden HT, Ullsperger M, Eichele T. Mal-adaptation of event-related EEG responses preceding performance errors. Frontiers in Human Neuroscience. 2010; 4

Habeck C, Moeller JR. Intrinsic functional-connectivity networks for diagnosis: just beautiful pictures? Brain Connectivity. 2011; 1:99–103. [PubMed: 22433005]

Harlow, LL.; Mulaik, SA.; Steiger, JH. What if there were no significance tests?. Lawrence Erlbaum; Mahwah, NJ: 1997.

Hengl, T. Visualisation of uncertainty using the HSI colour model: computations with colours; Paper presented at: Proceedings of the 7th International Conference on GeoComputation; Southhampton. 2003.

Hintze JL, Nelson RD. Violin plots: a box plot-density trace synergism. American Statistician. 1998; 52:181–184.

Jones, GE. How to lie with charts. 2nd. LaPuerta; 2006.

Kampstra P. Beanplot: A boxplot alternative for visual comparison of distributions. Journal of Statistical Software, Code Snippets. 2008; 28:1–9.

Keehner M, Mayberry L, Fischer MH. Different clues from different views: The role of image format in public perceptions of neuroimaging results. Psychonomic Bulletin & Review. 2011; 18:422–428. [PubMed: 21327381]

Kiehl KA, Laurens KR, Duty TL, Forster BB, Liddle PF. Neural sources involved in auditory target detection and novelty processing: an event-related fMRI study. Psychophysiology. 2001; 38:133–142. [PubMed: 11321614]

Lane DM, Sándor A. Designing better graphs by including distributional information and integrating words, numbers, and images. Psychological Methods. 2009; 14:239–257. [PubMed: 19719360]

Lieberman MD, Cunningham WA. Type I and Type II error concerns in fMRI research: re-balancing the scale. Social Cognitive and Affective Neuroscience. 2009; 4:423–428. [PubMed: 20035017]

McCabe DP, Castel AD. Seeing is believing: The effect of brain images on judgments of scientific reasoning. Cognition. 2008; 107:343–352. [PubMed: 17803985]

Tufte, E. The Visual Display of Quantitative Information. Graphics Press; Cheshire, CT: 2001.

Vaux DL. Error message. Nature. 2004; 428:799–799. [PubMed: 15103350]

Wainer H. Depicting error. American Statistician. 1996; 50:101–111.

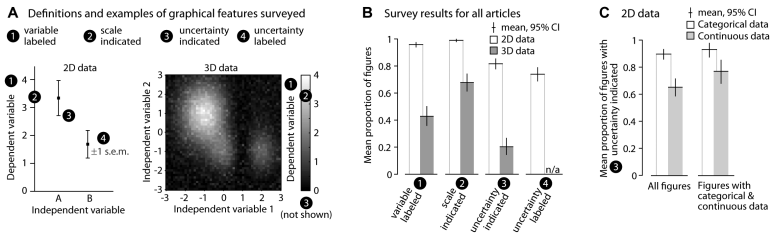


Figure 1. Survey results

(**A**) Definitions and examples of graphical features for 2D (left) and 3D (right) datasets. (**B**) Mean proportion of 2D (white) and 3D (dark gray) figures displaying each feature. Error bars denote 95% non-parametric confidence intervals (10,000 resamples). (**C**) Mean proportion of 2D figures indicating uncertainty, separated by categorical (white) and continuous data (light gray). Left panel considers all figures; right panel considers only figures with both categorical and continuous data.

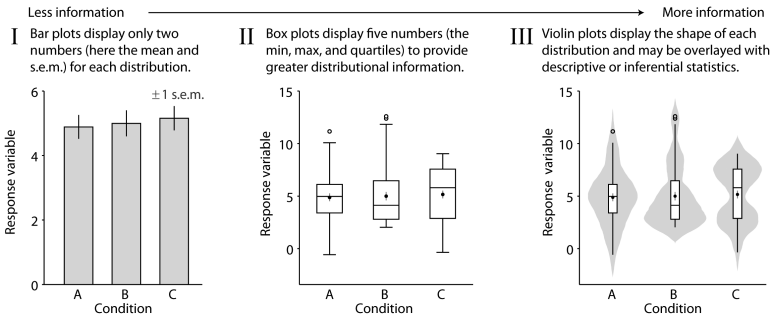


Figure 2. Comparison of graphical designs

The same synthetic data is summarized in a bar plot (I), box plot (II), and violin plot (III). Box plots in (II) and (III) also show the mean ± s.e.m. and are drawn with a maximum whisker length of 1.5 times the interquartile range. Data points (n=50 for each condition) were sampled from a normal distribution (condition A), a generalized χ^2 distribution with 2 degrees of freedom (B), and an equal mixture of two normal distributions with different means (C).

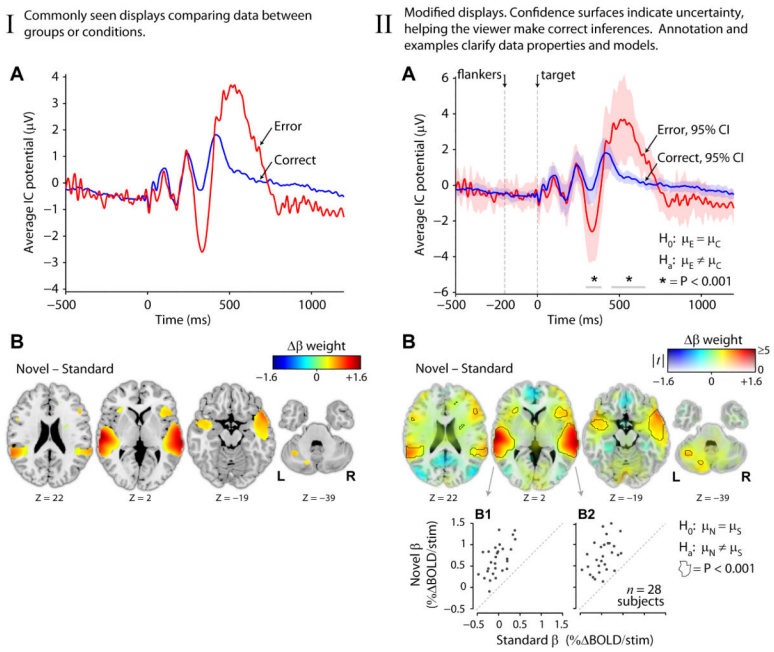
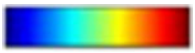
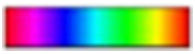


Figure 3. Conventional (I) and modified designs (II)

Captions describe panel II. (**A**) EEG flanker data. ERPs for error trials (red) and correct trials (blue) averaged over 10 subjects. Error bands are 95% non-parametric CIs (1000 bootstraps). Asterisks indicate significantly different ERPs at P<0.001 (nonparametric randomization test, 10,000 randomizations, implicit correction for multiple comparisons). (**B**) FMRI auditory oddball data. Axial slices show the difference between novel and standard beta weights averaged over 28 subjects. Beta difference is mapped to color hue; *t*-statistic magnitude is mapped to transparency. Contours denote significantly different betas at P<0.001 (two-tailed paired *t*-tests corrected with false discovery rate). (**B1**,**B2**) Scatter plots of standard versus novel betas for select regions. Beta weights are averaged over clusters of contiguous voxels passing significance (B1=2426 voxels; B2=1733 voxels). Dotted lines indicate y=x.

Table 1

When evaluating a figure for clarity and completeness, consider the following questions.

Questions	Examples/Suggestions
Design/Organization	
Is the display consistent with the model or hypothesis being tested?	• If data have been residualized or transformed for statistical analysis they should also be transformed in the graph.
Are there "empty dimensions" in the display that could be removed?	• If data are paired between conditions, the graph should reveal the pairwise differences rather than differences at the group level.
Does the display provide an honest and transparent portrayal of the data?	• A 3D pie chart for 2D categorical data • Extraneous colors that do not encode meaningful information • Hiding, smoothing, or modifying data has been avoided • Actual data points are emphasized over idealized models
Axes	
Are axes scales defined as linear, log, or radial?	
Does each axis label describe the variable and its units?	• For quantities with units: "Time to peak (ms)" • For arbitrary units (a.u.): "BOLD signal intensity (a.u.)"
Are axes limits appropriate for the data?	• For unitless quantities: "Spearman rank correlation"
Is the aspect ratio appropriate for the data?	• The graphic should not be bounded at zero if the data can take on both positive and negative values. • When x- and y-axes contrast the same variable under different conditions, the graphic should be square.
Color mapping	
Is a color bar provided?	
Is the color map sensible for the data type?	• Use  when data is bipolar, and map zero to green • Use  when data is unipolar, and map zero to black • Use  when data is circular, and map $-\pi$, $+\pi$ to red
Does the color bar axis indicate the quantity, units, and scale?	
Uncertainty	
Does the display indicate the uncertainty of estimated parameters?	
Is the type of error surface appropriate for the data?	• Standard deviations or prediction intervals are useful to describe variability in the population.
Are the units of uncertainty defined?	• Standard errors or confidence intervals are useful to make inferences about parameters estimated from a sample. • Parametric confidence intervals should only be used if data meet the assumptions of the underlying model.

Questions	Examples/Suggestions
	• "Error bands indicate non-parametric 95% confidence intervals of the median"
Color	
Are contrasting colors consistent with a natural interpretation?	• Red for increases, blue for decreases
Can features be discriminated when printed in grayscale?	• Group A  Group B 
Has red/green contrast been avoided to accommodate common forms of colorblindness?	
Annotation	
Information necessary to understand the display should be shown on the figure itself. Details & definitions may be relegated to the legend.	
Are all symbols defined, preferably by directly labeling objects?	
Is the directionality of a contrast between conditions obvious?	• "Patients – Controls"
Is the number of samples or independent experiments indicated?	• "Each point represents the mean over 23 subjects" • For a single test: “A repeated-measures ANOVA showed a significant effect of treatment (F[2, 10] = 12.53, P = 0.002)"
Are statistical procedures and criteria for significance described?	• For several tests: "Asterisks denote correlations different from zero (P < 0.01, two-tailed t-tests, Bonferroni corrected for 10 tests)."
Are uncommon abbreviations avoided or clearly defined?	
Are abbreviations consistent with those used in the text?	